Confidence Interval of Parameters of Some Well Known Distribution Functions

Thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

by

Barsa Priyadarsini Sarangi

Under the guidance of

Prof. Manas Ranjan Tripathy

Department of Mathematics
National Institute of Technology
Rourkela-769008
India

May 2014

Certificate

This is to certify that the project report entitled "**Confidence Interval of Parameters of Some Well Known Distribution Functions** " submitted by **Barsa Priyadarsini Sarangi** to the National Institute of Technology Rourkela, Orissa for the partial fulfilment of requirements for the degree of master of science in Mathematics is a bonafide record of review work carried out by her under my guidance. She has worked as a project student in this Institute for one year. In my opinion the work has reached the standard, fulfilling the requirements of the regulations related to the Master of Science degree. The contents of this project, in full or in parts, have not been submitted to any other institute or university for the award of any degree or diploma.

Place: NIT Rourkela
Date: May 2014

Dr. Manas Ranjan Tripathy
Assistant Professor
Department of Mathematics
NIT Rourkela-769008
India

Acknowledgements

It is my privilege to express my heartily feeling by dedicating this page to people who had been the living inspiration for me because of their humble co-operation and relentless encouragement. It is my immense pleasure to express my deep sense of gratitude and indebtedness to Prof. Manas Ranjan Tripathy, National Institute of Technology, Rourkela, my project guide for his ingenious guidance and constructive criticism throughout the entire period of my project work. His advice and involvement had helped a lot to give this project report, the present shape.

I would like to express the deepest appreciation to the faculty members of Department of Mathematics for their co-operation.

Lastly, I express my heartfelt devotion to my beloved parents and my friends for their blessings, love, affection and incessant inspiration.

Place: NIT Rourkela

Date: May 2014

(Barsa Priyadarsini Sarangi)

Roll No-412MA2082

Abstract

The thesis addresses the study of some basic results used in statistics and estimation of parameters. Here we mainly focus on interval estimation. Here we have presented confidence interval of parameters of some well known distribution functions and bayesian interval estimation of some distribution functions. Here we briefly discuss about Theory of Estimation that is about Characteristics of good estimators, Methods of estimation, Confidence interval and Bayesian interval. We also deals with certain results which are very much useful for this thesis and also we discuss the importance of this in real life problem, with appropriate examples.

Contents

1 Introduction **1**

2 Some Definitions and Basic Results **5**

2.1 Some Basic Definitions 5

2.2 Basic Concepts for Estimation of Parameters 8

2.3 Characteristics of Estimator 9

2.4 Method of Estimation 12

2.4.1 Method of Maximum Likelihood Estimation 12

2.4.2 Method of Moments 14

2.4.3 Method of Minimum Variance 16

2.4.4 Method of Least Squares 16

2.5 Interval Estimation 17

2.5.1 Bayesian Estimation 18

3 Finding Confidence Interval of Parameters of Normal and Binomial Distribution **21**

3.1 Confidence Interval for Parameters of the Normal Distribution 22

3.1.1 Confidence Interval for mean μ of the Normal Distribution with known variance σ^2 22

3.1.2 Confidence interval for mean μ of the Normal distribution with unknown variance σ^2 24

3.1.3 Confidence interval for the variance σ^2 of the Normal distribution . 26

3.2 Binomial confidence intervals 28

4 Bayesian Interval Estimation **31**

4.1 Exponential Distributions 31

4.2 Normal Distribution 34

4.3 Weibull Distribution 37

Conclusions and Scope of FutureWorks **39**

Bibliography **41**

Chapter 1

Introduction

One of the important research area in the field of statistics is the statistical inference. Statistical inference is the process by which information from the sample data is used to draw conclusions about the population from which the sample was selected. Statistical inference can be classified into two types one is the estimation of the parameters and the other being the hypothesis testing about the parameter involved in a distribution function. Further estimation refers to two types one is the point estimation and the other is the confidence interval. The theory of estimation was founded by Prof. R. A. Fisher in a series of fundamental papers round about 1980. A point estimate of a parameter is a number, which is computed from a given sample and serves as an approximation of the unknown exact values of the parameter. The estimation procedure requires an estimation rule that can be applied to data which have been collected according to some acceptable procedure. An interval estimation is an interval (confidence interval) obtained from a sample. Estimation of parameters is of great practical importance in many applications.

Suppose a random sample of n observations from the population is available, the values of the observations being $x_1, x_2, ..., x_n$. So we can calculate the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, and use this as an estimate of μ. Here $\bar{x}$ is a point estimate of μ. Now of course the actual numerical values we obtain for $\bar{x}$ depends on which particular members of the population fall in the sample. We know that we shall obtain a different value of $\bar{x}$ from a different sample. So our point estimate will vary from sample to sample, and so we cannot say how near to μ a point estimate from a single sample will be. What we can say that if $\bar{X}$ provides an unbiased estimator of μ, (that is, $E[\bar{X}] = \mu$), so that if many samples each of size n, are taken from the same population, the mean of the sampling distribution of $\bar{X}$ will be μ. In practice, we usually have a single sample of n observations, from which we have calculated the mean $\bar{x}$, and we want to extract as much information as possible from it. For example a manufacturer of a food product which is packed automatically has to set his machinery to give a correct mean weight μ. After servicing or resetting the machine, he wishes immediately to take a sample of packets from his production line and check that mean weight μ. He does not want to wait for many samples and many values of $\bar{x}$, he want to find about μ from this one sample. Another example can be, suppose that a doctor is giving a drug to a patient in order to control the level of the patient's blood sugar. The doctor knows that in order to avoid danger to the patient the true blood sugar level has to be within certain limits. He obviously cannot take a very large number of blood samples, he must be satisfied with one sample of n observations, n being perhaps quite a small number, and on the basis of this he must decide whether the true mean blood sugar level in the whole bloodstream is within the acceptable limits.

In this project work we shall make it possible to give an interval estimate for parameter using random sample. We find upper and lower limits for parameters, and hence an interval which is very likely to contain that parameter, because we can say just how likely our interval is to contain the true values of parameters. This type of estimate gives more information than the previous point estimate. When estimating proportions, we have similar problem. So far, a true proportion π in a population has been estimated by calculating the sample proportion p, this gives us a point estimate of π. Suppose, for example, that we are carrying out a survey to see what proportion of residents in a village would use a local bus service if it ran at a certain time at 8.30 a.m. We can either ask every resident or, much more practically, we can base an estimate on the proportion found in a sample. Obviously here too it will be very useful to be able to give limits that are very likely to contain π. These would enable us to know whether the bus is likely to be nearly empty or very full.

Let us consider one more example to get a better idea how confidence interval behaves better than point estimation. In the production of pipe fitting it might be more important to specify an interval believe to contain the mean of the inside diameters rather than have a point estimate of the mean inside diameter. That is we may be more interested in the mean being within some acceptable bounds that in having some estimate of its actual value. In this thesis we try to give an interval estimation for parameters of some distribution functions.

Bayesian confidence interval differs from the classical or frequenting confidence interval due to difference in the interpretation of the parameter θ. To construct a $(1-\alpha)$ credible interval for θ we take a function $g(\underline{x}, \theta)$ of the parameter θ and the observation $\underline{x}$ such that the distribution of $g(\underline{x}, \theta)$ does not dependent on any unknown parameter.

For the classical confidence interval θ is a constant and the end points are random variables, and in Bayesian credible interval, θ is the random variable and the end points are fixed. The rest of the chapters are organized as follows. In Chapter 2, we discuss some basic results related to interval estimation. In Chapter 3, we discuss confidence interval for parameters of like μ and σ of Normal distribution. And also we discussed confidence interval for parameters of Binomial distribution. Finally, in Chapter 4, we study about finding bayesian interval for parameters of certain distribution function like Exponential, Normal, and Weibull.

Chapter 2

Some Definitions and Basic Results

In this chapter we discuss some basic definitions and results which are very much useful for the development of our consequence chapters.

2.1 Some Basic Definitions

Definition 2.1 (Random experiment) *It is an experiment in which outcomes are known but the performance of the experiment is unknown. This experiment can be repeated under identical conditions.*

Definition 2.2 (Sample space) *Sample space of an random experiment is a pair* (Ω, S), *where* Ω *is the set of all possible outcomes of the experiment and* S *is the* σ *field of a subset of* Ω.

Definition 2.3 (Event) *A subset of a sample space* Ω *in which a statistician is interested is known as event.*

Definition 2.4 (Probability measure) *Let* (Ω, S) *be a sample space. A set function* P

defined on S is called probability or simply probability measure if it satisfies the following condition,

(*i*) $P(A) \geq 0, \forall\ A \in S$.

(*ii*)$P(\Omega) = 1$.

(*iii*) *Let* A_j , $A_j \in S, j = 1, 2, \ldots$ *be a disjoint sequence of sets. That is* $A_j \bigcap A_k = \emptyset$ *for* $j \neq k$. *Then*

$$P\Big(\bigcup_{j=1}^{\infty}\Big) = \sum_{j=1}^{\infty} P\Big(A_j\Big).$$

Definition 2.5 (Ramdom variable) *Random variable X is a function from sample Ω to the set of real numbers such that the inverse image of a Borel set in $\mathbb{R}$, under X is an event. That is, $X : \Omega \longrightarrow \mathbb{R}$ such that $X^{-1}(-\infty, a] \in S,\ a \in \mathbb{R}$.*

Definition 2.6 (Distribution function) *Let X be a random variable defined on (Ω, S, P). Define a function F on $(-\infty, \infty)$ by $F(x) = P\{w : X(w) \leq x\}$, for all $x \in \mathbb{R}$. F is nondecreasing, right continues and $F(-\infty) = 0$, $F(\infty) = 1$. Then the function F is called the distribution function of the random variable X.*

Depending upon the nature of sample space we may categorize the random variables as discrete type and continuous type.

Definition 2.7 (Discrete random variable) *A random variable X which is define on (Ω, X, P) is called discrete type if there exist a countable set $E \subseteq \mathbb{R}$, such that $P\{X \in E\} = 1$.*

Definition 2.8 (Probability mass function) *The collection of number p_i which satisfies $P\{X = x_i\} = p_i \geq 0$ for all i and $\sum_{i=1}^{\infty} p_i = 1$ is known as probability mass function of RV X.*

Definition 2.9 (Continuous random variable) *Let X be a random variable define on (Ω, S, P) with distribution function F. Then X is said to be continuous if F is absolutely continuous, that is, if there exists a non negative function $f(x)$ such that for every real number x,*

$$F(x) = \int_{-\infty}^{x} f(t)dt.$$

Definition 2.10 (Probability density function) *f is called the density function of the random variable X if it satisfies $f \geq 0$ and $\int_{-\infty}^{\infty} f(t)dt = 1$.*

Definition 2.11 (Marginal distribution function) *Let (X, Y) be two dimensional random variable with joint distribution function $F(x, y)$, then the marginal distribution of X is*

$$\begin{aligned} F_1(x) = F_X(x) &= \sum_{x_i \leq x} p_{i.} = \sum_{x_i \leq x} P(X = x_i), \text{ if } (X, Y) \text{ is discrete,} \\ &= \int_{-\infty}^{x} f_1(t)dt, \text{ if } f(X, Y) \text{ is continuous.} \end{aligned}$$

Similarly the marginal distribution on Y is denoted by

$$\begin{aligned} F_2(y) = F_Y(y) &= \sum_{y_i \leq y} p_{.j} = \sum_{y_i \leq y} P(Y = y_i), \text{ if } (X, Y) \text{ is discrete,} \\ &= \int_{-\infty}^{y} f_2(t)dt, \text{ if } f(X, Y) \text{ is continuous.} \end{aligned}$$

Definition 2.12 (Mathematical expectation) ***Discrete case*** *If X be a discrete random variable having PMF $P_k = P(X = x_k)$ then, we can say the mathematical expection of X exist and write $E(X) = \sum_{k=1}^{\infty} x_k P_k$ provided $\sum_{k=1}^{\infty} |x_k| P_k < \infty$.*
Continuous case *If X be a Continuous random variable having density function f then, we can say the mathematical expection of X exist and write $E(X) = \int_{-\infty}^{\infty} x f(x)dx$ provided $\int_{-\infty}^{\infty} |x| f(x)dx$.*

Definition 2.13 (Variance) *If $E(X^2)$ exists, the variance is defined by $\sigma^2 = Var(x) = E(X - \mu)^2$. The quantity σ is called the standard deviation of X and $\sigma^2 = \mu_2 = E(X^2) - (E(X))^2$.*

Next, we discuss some basic results and terminologies related to the estimation of parameters of a distribution function.

2.2 Basic Concepts for Estimation of Parameters

Suppose $X_1, X_2, \ldots, X_n$ are collected from a population which has a distribution function $F_\theta(x)$. This is normally a family of distribution functions as for each value of the parameter θ we have a distribution F. Here θ may lie in some set say Θ (parameter space), is the unknown parameter associated with the distribution function F.

Definition 2.14 (Estimation of parameters) *Suppose $F_\theta(x)$, $\theta \in \Theta$ be a family of distribution functions and θ is taken to be unknown. Here we estimate the unknown parameter θ with the help of samples.*

Definition 2.15 (Ramdom sample) *A random sample of size n corresponding to the random variable X is the collection of n independent random variable $X_1, X_2, \ldots, X_n$ such that X_k, $k = 1, 2, \ldots, n$ is identically distributed with X.*

Definition 2.16 (Parameter space) *Let X be a random variable defined on a sample space Ω having probability mass function $f(x, \theta)$. Here θ is unknown and takes the values on a set called as parameter space Θ.*

Definition 2.17 (Statistics) *A function of the random variables comprising a random sample is called a statistic.*

Definition 2.18 (Estimator) *If a statistic $T(X_1, X_2, \ldots, X_n)$ is used to estimate an unknown parameter θ of a distribution, then it is called an estimator and a particular value of the estimator say $T_n(X_1, X_2, \ldots, X_n)$ is called an estimate of θ.*

The process of estimating an unknown parameter is known as estimation.

Next we discuss some criteria that should be satisfied by a good estimator.

2.3 Characteristics of Estimator

Various statistical properties of an estimators can be used to decide which estimator is most appropriate in a given situation.

Definition 2.19 (Consistency) *Let $X_1, X_2, \ldots, X_n$ be a sequence of iid random variables with common distribution function F_θ, $\theta \in \Theta$. A sequence of point estimators $T_n(X_1, X_2, \ldots, X_n) = T_n$ will be called consistent for $\psi(\theta)$ if T_n converges to $\psi(\theta)$ in probability, that is,*

$$P(|T_n - \psi(\theta)| > \epsilon) \to 0, \ \text{as } n \to \infty,$$

where ϵ is a very small arbitrary positive number.

Example 2.1 *If $X_1, X_2, \ldots, X_n$ is a random sample from a papulation with finite mean $E(X_i) = \mu < \infty$, then by Khinchen's weak law of large number we have,*

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E(X_i) = \mu, \text{as } n \to \infty,$$

where sample mean $(\overline{X}_n)$ is always a consistent estimator of the population mean μ.

Below we discuss some of the consequences of the above definition.

Remark 2.1 *If T_n is a consistent estimator of θ and $\psi(\theta)$ is a continuous function of θ, then $\psi(T_n)$ is a consistent estimator of $\psi(\theta)$.*

Remark 2.2 *If T_n is a sequence of consistent estimators such that $E[T_n] \to \psi(\theta)$ and $Var[T_n] \to 0$ as $n \to \infty$, then T_n is a consistent estimator of $\psi(\theta)$.*

Definition 2.20 (Unbaisedness) *Unbaisedness is a property associated with finite n. A statistic $T_n = T(x_1, x_2, \ldots, x_n)$ is said to be an unbaised estimator of $\gamma(\theta)$ if $E(T_n) = \gamma(\theta)$, for all $\theta \in \Theta$.*

Remark 2.3 *If $E(T_n) > \theta$, T_n is said to be positively biased and if $E(T_n) < \theta$, T_n is said to be negatively biased. The amount of bias $b(\theta)$ being given by $b(\theta) = E(T_n) - \gamma(\theta)$, $\theta \in \Theta$.*

Example 2.2 *If $X_1, X_2, \ldots, X_n$ is a random sample from a normal population $N(\mu, 1)$. Show that $T = \frac{1}{n}\sum_{i=1}^{n} X_i^2$is an unbiased estimator of $\mu^2 + 1$.*
Solution *Here given that $E(X_i) = \mu$, $V(X_i) = 1$, for all $i = 1, 2, \ldots, n$.*
We know that

$$\begin{aligned} Var(X_i) &= E(X_i^2) - (E(X_i))^2 \\ E(X_i^2) &= Var(X_i) + (E(X_i))^2 \\ &= 1 + \mu^2 \end{aligned}$$

$$\begin{aligned} E(T) &= E\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} E(X_i^2) \\ &= \frac{1}{n}\sum_{i=1}^{n} 1 + \mu^2 \\ &= \frac{1}{n} n(1 + \mu^2) \\ &= 1 + \mu^2 \end{aligned}$$

$E(T) \in N(\mu, 1)$, Hence T is an unbiased estimator of 1+μ^2.

Definition 2.21 (Efficiency) *If T_1 is the most efficient estimator with V_1 and T_2 is any other estimator with variance V_2, then efficiency E of T_2 is $E = \frac{V_1}{V_2}$.*

If $T, T_1, T_2, \ldots, T_n$ are all estimator of $\gamma(\theta)$ and $Var(T)$ is minimum, then the efficiency of E_i of T_i is $E_i = \frac{Var(T)}{Var(T_i)}$, $i = 1, 2, \ldots, n$. So, $E_i \leq 1, i = 1, 2, \ldots, n$.

Remark 2.4 *If in a class of consistent estimator for a parameter, there exists one whose sampling variance is less than that of any such estimator, it is called the Most efficient estimator.*

Minimum Variance Unbiased(MVU) Estimator: If a statistic $T = T(X_1, X_2, \ldots, X_n)$ based on sample of size n is such that
(i) T is unbiased for $\gamma(\theta)$, for all $\theta \in \Theta$,
(ii) It has the smallest variance among the class of all unbiased estimator of $\gamma(\theta)$, then T is called the minimum variance unbiased estimator (MVUE) of $\gamma(\theta)$. That is, T is MVUE of $\gamma(\theta)$ if
$E_\theta(T) = \gamma(\theta)$, for all $\theta \in \Theta$ and $Var_\theta(T) \leq Var_\theta(T_1)$, for all $\theta \in \Theta$,
where T_1 is any other unbiased estimator of $\gamma(\theta)$.

Definition 2.22 (Sufficiency) *An estimator is said to be sufficient for a parameter, if it contains all the information in the sample regarding the parameter. If $T = T(x_1, x_2, \ldots, x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2, \ldots, x_n$ of size n from the population with density $f(x, \theta)$ such that the conditional distribution of $x_1, x_2, \ldots, x_n$ given T is independent of θ, then T is sufficient estimator for θ.*

Next, we discuss a technique known as "Fisher-Neymann factorization criterion" to determine sufficient statistics for a given distribution.

Theorem 2.1 (Fisher-Neymann factorization criteria) *A statistic $T = t(X)$ is a sufficient statistic for the parameter θ if and only if the joint probability distribution or density of the random sample can be expressed in the form:*

$$f(x_1, x_2, \ldots, x_n; \theta) = g_\theta(t(x)) \times h(x_1, x_2, \ldots, x_n),$$

where $g_\theta(t(x))$ depends on θ and x and $h(x_1, x_2, \ldots, x_n)$ does not depend on θ.

Example 2.3 *Let $x_1, x_2, \ldots, x_n$ be a random sample from the bernoulli population with parameter p, $0 < p < 1$, That is,*

$$x_i = 1, \text{ with probability } p,$$

$$= \quad 0, \ \textit{with probability } q = 1 - p.$$

Then $T = t(x_1, x_2, \ldots, x_n) = x_1 + x_2 + \ldots + x_n \sim B(n, p)$
$\therefore P(T = k) = \binom{n}{k} p^k (1-p)^{n-k}$.
The conditional distribution of $x_1, x_2, \ldots, x_n$ *given* T *is*

$$\begin{aligned} P\left[x_1 \cap x_2 \cap \ldots \cap x_n | T = k\right] &= \frac{P\left[x_1 \cap x_2 \cap \ldots \cap x_n \cap T = k\right]}{P(T = k)} \\ &= \frac{p^k (1-p)^{n-k}}{\binom{n}{k} p^k (1-p)^{n-k}} = \frac{1}{\binom{n}{k}}, \\ &= 0, \quad if \quad \sum_{i=1}^{n} x_i \neq k. \end{aligned}$$

So, here it does not depend on p*. So* $T = \sum_{i=1}^{n} x_i$ *is sufficient for* p*.*

2.4 Method of Estimation

Normally there are four different approaches for obtaining a point estimator for unknown parameter θ. Namely classical method and decision theoretic approach. Now we outline some of the most important methods for obtaining point estimators. Most commonly used methods under classical estimation are as follows.

2.4.1 Method of Maximum Likelihood Estimation

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a population with density function $f(x, \theta)$. Then the likelihood function of the sample values $x_1, x_2, \ldots, x_n$ ($L = L(\theta)$) is their joint density function.
$L = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$.

The principle of maximum likelihood consist in finding an estimator for the unknown parameter $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$ which maximizes the likelihood function $L(\theta)$ for variation in parameter.
That is, we wish to find $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$, so that $L(\hat{\theta}) > L(\theta)$, for all $\theta \in \Theta$.

That is, $L(\hat{\theta}) = Sup\ L(\theta)$, for all $\theta \in \Theta$.

Thus if there exist a function $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ of the sample value which maximizes L for variation in θ, then $\hat{\theta}$ is to be taken as an estimator of θ. $\hat{\theta}$ is called maximum likelihood estimator (MLE). Thus $\hat{\theta}$ is the estimator if

$$\frac{\partial L}{\partial \theta} = 0 \ \ and \ \ \frac{\partial^2 L}{\partial \theta^2} < 0. \tag{2.1}$$

Since $L > 0$, and $\log L$ is a non zero decreasing function of L; L and $\log L$ attain their extreme value at the same value of $\hat{\theta}$. Then equation (2.1) will be

$$\frac{1}{L} \cdot \frac{\partial L}{\partial \theta} = 0 \Rightarrow \frac{\partial \log L}{\partial \theta} = 0. \tag{2.2}$$

If θ is a vector values parameter $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$, then

$$\frac{\partial}{\partial \theta_i} \log L = \frac{\partial}{\partial \theta_i} \log L(\theta_1, \theta_2, \ldots, \theta_k) = 0, \quad i = 1, 2, \ldots, k. \tag{2.3}$$

Equation (2.2) and (2.3) are the likelihood equations for estimating the parameters.

Theorem 2.2 (Cramer -Rao Theorem) *With probability approaching unity as $n \to \infty$, the likelihood equation $\frac{\partial}{\partial \theta} \log L$, has a solution which converges in probability to the true value θ_0 , That is, MLE's are consistent.*

Example 2.4 *Find maximum likelihood estimates for $\theta_1 = \mu$ and $\theta_2 = \sigma$ in the case of the Normal distribution.*

Solution *The Normal distribution function is*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

The likelihood function is

$$\begin{aligned} l &= (f(x_i))^n \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \end{aligned}$$

Taking logarithms, we have,

$$L = \log l = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

First we have to take, $\frac{\partial}{\partial\mu}(L) = 0$

$$\begin{aligned}
\frac{\partial L}{\partial \mu} = -\frac{\partial}{\partial \mu}\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) &= 0 \\
\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) &= 0 \\
\Rightarrow \sum_{i=1}^{n} x_i - n\mu &= 0
\end{aligned}$$

The solution is the desired estimate $\hat{\mu}$ *for* μ,

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

Then, we have to take $\frac{\partial}{\partial\sigma}(L) = 0$

$$\begin{aligned}
\Rightarrow -\frac{n}{2}\frac{1}{\sigma^2}2\sigma + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 &= 0 \\
\Rightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\mu)^2 &= 0 \\
\Rightarrow -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= 0 \\
\Rightarrow \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\bar{x})^2 &= n \\
\Rightarrow \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
\therefore \hat{\sigma^2} &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.
\end{aligned}$$

2.4.2 Method of Moments

If $f(x, \theta_1, \theta_2, \dots, \theta_k)$ be the density function of the parent population with k parameters $(\theta_1, \theta_2, \dots, \theta_k)$. If μ_r' denotes the r^{th} moment about origin, Then,

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x;\theta_1,\theta_2,...,\theta_k)dx,\ r = 1, 2, \dots, k$$

In general $\mu_1', \mu_2', \mu_3', \ldots, \mu_k'$ will be the function of the parameter $(\theta_1, \theta_2, \ldots, \theta_k)$.
Let $x_i, i = 1, 2, \ldots, n$ be the random sample of size n from the given population. The method of moments consists in solving the k equation for $\theta_1, \theta_2, \ldots, \theta_k$ in terms of $\mu_1', \mu_2', \mu_3', \ldots, \mu_k'$ and then replacing these moments μ_r', $r = 1, 2, \ldots, k$ by the sample moments.
That is, $\hat{\theta}_i = \theta_i(\hat{\mu_1}', \hat{\mu_2}', \ldots, \hat{\mu_k}') = \theta_i(m_1', m_2', m_3', \ldots, m_k')$, $i = 1, 2, \ldots, k$,
where m_i is the i^{th} moment about origin in the sample $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ are the required estimators of $\theta_1, \theta_2, \ldots, \theta_k$ respectively.

Example 2.5 *Estimate α and β in the case of pearson's type III disrtibution by the method of moments.*

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \qquad 0 \leq x \leq \infty.$$

Solution *We have,*

$$\begin{aligned}
\mu_r' &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^r x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{(r+\alpha-1)} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha + r)}{\beta^{\alpha+r}} = \frac{\Gamma(\alpha + r)}{\Gamma(\alpha)\beta^r}
\end{aligned}$$

$$\mu_1' = \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)\beta} = \frac{\alpha!}{(\alpha - 1)!\beta} = \frac{\alpha}{\beta}.$$

$$\mu_2' = \frac{\Gamma(\alpha + 2)}{\Gamma(\alpha)\beta^2} = \frac{(\alpha + 1)\alpha}{\beta^2}.$$

$$\begin{aligned}
\frac{\mu_2'}{\mu_1'^2} &= \frac{\alpha + 1}{\alpha} = \frac{1}{\alpha} + 1 \\
\Rightarrow \frac{1}{\alpha} &= \frac{\mu_2'}{\mu_1'^2} - 1 = \frac{\mu_2' - \mu_1'^2}{\mu_1'^2} \\
\Rightarrow \alpha &= \frac{\mu_1'^2}{\mu_2' - \mu_1'^2}.
\end{aligned}$$

$$\beta = \frac{\alpha}{\mu_1'} = \frac{\frac{\mu_1'^2}{\mu_2' - \mu_1'^2}}{\mu_1'} = \frac{\mu_1'}{\mu_2' - \mu_1'^2}.$$

Hence,

$$\hat{\alpha} = \frac{m_1'^2}{m_2' - m_1'^2}, \quad \hat{\beta} = \frac{m_1'}{m_2' - m_1'^2}.$$

Now let us discuss some decision theoretic approach for point estimation of the parameters. Most commonly used methods under decision theoretic approach are as follows.

2.4.3 Method of Minimum Variance

Here the estimator is (i) unbiased and (ii) minimum variance.
If $L = \prod_{i=1}^{n} f(x_i, \theta)$ is the likelihood function of a random sample of n observations $x_1, x_2, \ldots, x_n$ from a population with probability function $f(x, \theta)$, then the problem is to find a statistic $T = t(x_1, x_2, \ldots, x_n)$, such that

$$E(T) = \int_{-\infty}^{\infty} t.Ldx = \gamma(\theta) \Rightarrow \int_{-\infty}^{\infty} [t - \gamma(\theta)]\, Ldx = 0 \tag{2.4}$$

$$V(T) = \int_{-\infty}^{\infty} [t - E(T)]^2\, Ldx = \int_{-\infty}^{\infty} [t - \gamma(\theta)]^2\, Ldx \tag{2.5}$$

is minimum where $\int_{-\infty}^{\infty} dx$ represent the n-fold integration

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dx_1 dx_2 dx_3 \ldots dx_n.$$

2.4.4 Method of Least Squares

The principle of least square is used to fit a curve of the form : $y = f(x, a_0, a_1, \ldots, a_n)$ where $a_i's$ are unknown parameters to a set of n sample observations (x_i, y_i), $i = 1, 2, \ldots, n$ from a bivariate population.
It consists in minimizing the sum of squares of residuals

$$E = \sum_{i=1}^{n} [y_i - f(x_i, a_0, a_1, \ldots, a_n)]^2$$

Subject to variations in $a_0, a_1, \ldots, a_n$.
The normal equation for estimating $a_0, a_1, \ldots, a_n$ are given by

$$\frac{\partial E}{\partial a_i} = 0, \quad i = 1, 2, \ldots, n.$$

2.5 Interval Estimation

Rather than finding a point estimate θ^* of an unknown parameter θ, it is sometimes more beneficial to locate an interval (θ_1^*, θ_2^*) which is belived to contain θ.

Definition 2.23 (Confidence Interval) *If $(\hat{\theta}_1, \hat{\theta}_2)$ is an interval estimator for which*

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha$$

and θ_1^* and θ_2^* are estimates resulting from a particular set of sample values, then the interval (θ_1^*, θ_2^*) is called a $(1-\alpha)100\%$ confidence interval, $(1-\alpha)$ is called the confidence coefficient and the endpoints θ_1^* and θ_2^* are called the lower and upper confidence limits respectively.

Definition 2.24 (Confidence Co-efficient) *Let $\theta \in \Theta \subseteq \mathbb{R}$ and $0 < \alpha < 1$. A function $\underline{\theta}(X)$ satisfying $P_\theta\{\underline{\theta}(X) \leq \theta\} \geq 1 - \alpha$, for all θ, is called a lower confidence bound for θ at confidence level $1 - \alpha$. The quantity*

$$inf_{\theta \in \Theta} P_\theta\{\underline{\theta}(X) \leq \theta\}$$

is called the confidence co-efficient.

Definition 2.25 (Uniformally most accurate (UMA) lower confidence bound) *A function $\underline{\theta}$ that minimizes*

$$P_\theta\{\underline{\theta}(X) \leq \theta'\}, \ for\ all\ \theta' < \Theta,$$

Subjected to

$$P_\theta\{\underline{\theta}(X) \leq \theta\} \geq 1 - \alpha, \ for\ all\ \theta,$$

is known as a uniformally most accurate (UMA) lower confidence bound for θ at confidence level $1 - \alpha$.

Definition 2.26 (Error Bounds) *To construct a confidence interval for a single unknown population mean μ, where the population standard deviation is known, we need $\bar{x}$ as an estimate for μ and we need the margin of error. Here, the margin of error is called the error bound(EBM) for a population mean. The sample mean $\bar{x}$ is the point estimate of the unknown population mean μ.*

Then the confidence interval estimate will have the form (point estimate - error bound, point estimate + error bound). That is in symbols, $(\bar{x} - EBM, \bar{x} + EBM)$. The margin of error depends on the confidence level.

2.5.1 Bayesian Estimation

In Bayesian Principle the unknown parameter θ which is treated as random variable assumes a probability distribution known as a prior of θ denoted by $\Pi(\theta)$.

To start the estimation of parameters we have the prior information about the unknown parameter θ. Different types of prior are discussed below.

(a) Noninformative Prior A pdf $\Pi(\theta)$ is said to be a noninformative prior if it contains no information about θ. Some simple examples of noninformative priors are $\Pi(\theta) = 1$, $\Pi(\theta) = \frac{1}{\theta}$.

(b) Jeffreys' invariant prior Jeffreys suggested a general rule for choosing the noninformative prior $\Pi(\theta)$. Where,

$$\Pi(\underline{\theta}) \propto \sqrt{I(\underline{\theta})}$$

where $\underline{\theta}$ vector valued parameter, and

$$I(\underline{\theta}) = -E\Big[\frac{\partial^2 \log f(x|\underline{\theta})}{\partial\theta_i \partial\theta_j}\Big] \tag{2.6}$$

where $I(\underline{\theta})$ is Fisher information matrix.

Definition 2.27 (posterior distribution) *The posterior distribution of θ given $X = x$ is obtained by dividing the joint density of θ and X by the marginal distribution of X. Mathematically*

$$\frac{\Pi(\theta)f_\theta(x)}{\int_\Theta \Pi(\theta)f_\theta(x)d\theta},$$

where Θ is the parameter space.

Remark 2.5 *Let Ω is the parameter space of θ, $g(\theta)$ is the prior distribution and $\Pi(\theta|x)$ is the posterior distribution of θ then,*

$$\Pi(\theta|x) = \frac{g(\theta)f(x|\theta)}{h(x)} = Cg(\theta)f(x|\theta), \tag{2.7}$$

where $C^{-1} = \int_\Omega \Pi(\theta|x)d\theta$.

Let $\underline{x} = (x_1, x_2, \ldots, x_n)$ be a random sample from the density function $f(x|\theta)$, then from (2.7) $\Pi(\theta|\underline{x}) = Cg(\theta)L(\underline{x}|\theta)$, *where $L(\underline{x}|\theta)$ is the corresponding likelihood function of the data $\underline{x}$.*

Definition 2.28 (Credible interval) *The interval $[\theta_1, \theta_2]$ based on the posterior distribution of θ is called a "Credible interval", which contain a certain fraction of one degree of belief such that*

$$1 - \alpha = P(\theta_1 < \theta < \theta_2) = \int_{\theta_1}^{\theta_2} \Pi\left(\theta|\underline{x}\right) d\theta, \tag{2.8}$$

where $\Pi\left(\theta|x\right)$ is the posterior distribution of θ, such an interval $[\theta_1, \theta_2]$ is known as the $(1-\alpha)$ credible interval of θ.

Definition 2.29 (Highest posterior density or HPD interval) *The interval I which simultaneously satisfies* (2.7) *and* (2.8) *is called the shortest $(1-\alpha)$ credible interval.*

An interval I which satisfies the following,

(i) that the interval be shortest possible ,and

(ii) $\Pi\left(\theta|\underline{x}\right), \theta \in I > \Pi\left(\theta|\underline{x}\right), \theta \notin I$,

That is, the posterior density at every point inside be $>$ the posterior density at every point outside it (the interval includes more probable values of θ and exclude less probable values of the parameter), is called the Highest posterior density or HPD interval.

Chapter 3

Finding Confidence Interval of Parameters of Normal and Binomial Distribution

Confidence interval for an unknown parameter θ of some distribution are interval $\theta_1 \leq \theta \leq \theta_2$ that contained θ, not with certainty but with a high probability γ which we can choose (90%, 95%,...), such an interval is calculated from a sample. Instead of witting $\theta_1 \leq \theta \leq \theta_2$, we can write this as $CONF_\gamma\{\theta_1 \leq \theta \leq \theta_2\}$, where γ is a confidence level. θ_1 & θ_2 are the lower and upper confidence limits. They depend on γ. The larger we choose γ, the smaller the error probability $1-\gamma$, and longer the confidence interval. If $\gamma \to 1$, then its length goes to infinity. θ_1 and θ_2 are can be calculated from a sample $x_1, x_2, \ldots, x_n$, where these are n observation of random variable X. We take $x_1, x_2, \ldots, x_n$ as single observations of n random variables $X_1, X_2, \ldots, X_n$. Then $\theta_1 = \theta_1(x_1, x_2, \ldots, x_n)$ and $\theta_2 = \theta_2(x_1, x_2, \ldots, x_n)$ are observed value of two random variables $\Theta_1 = \Theta_1(X_1, X_2, \ldots, X_n)$ and $\Theta_2 = \Theta_2(X_1, X_2, \ldots, X_n)$.
Then, $P\{\Theta_1 \leq \theta \leq \Theta_2\} = \gamma$.

3.1 Confidence Interval for Parameters of the Normal Distribution

3.1.1 Confidence Interval for mean μ of the Normal Distribution with known variance σ^2

$1^{st} Step$: Choose a confidence level $\gamma(95\%, 99\%, \ldots)$.
$2^{nd} Step$: Determine the corresponding c,

$\gamma:$ 0.90 0.95 0.99 0.999

$c:$ 1.645 1.960 2.576 3.291

$3^{rd} Step$: Compute the mean $\bar{x}$ of the sample $x_1, x_2, ..., x_n$.
$4^{th} Step$: Compute $k = c\sigma/\sqrt{n}$, The confidence interval for μ is

$$CONF_{\gamma}\{\bar{x} - k \leq \mu \leq \bar{x} + k\}. \tag{3.1}$$

Example 3.1 *Determine a* 95% *confidence interval for the mean of normal distribution with variance* $\sigma^2 = 9$, *using a sample of* $n = 100$ *values with mean* $\bar{x} = 5$.
Solution
$1^{st} Step$: 95% *confidence interval, So* $\gamma = 0.95$
$2^{nd} Step$: *The corresponding* $c = 1.960$
$3^{rd} Step$: $\bar{x} = 5$ *(given)*
$4^{th} Step$: $k = c\sigma/\sqrt{n} = 1.960 \times 3/\sqrt{100} = 0.588$

Hence, $\bar{x} - k = 5 - 0.588 = 4.412$, $\bar{x} + k = 5 + 0.588 = 5.588$

The confidence interval is,

$$CONF_{0.95}\{4.412 \leq \mu \leq 5.588\}.$$

Theorem 3.1 (Sum of independent normal random variables) *Let* $(X_1, X_2, \ldots, X_n)$ *be the independent random variable each of which has mean* μ *and variance* σ^2. *Then the following holds,*
(*a*) *The sum* $(X_1 + X_2 + \ldots + X_n)$ *is normal with mean* $n\mu$ *and variance* $n\sigma^2$.

(b) The following random variable $\bar{X}$ is normal with mean μ and variance $\frac{\sigma^2}{n}$.

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) \tag{3.2}$$

(c) The following random variable Z is normal with mean 0 and variance 1.

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}. \tag{3.3}$$

proof(a)

$$\begin{aligned} Mean &= E(X_1 + X_2 + \ldots + X_n) \\ &= EX_1 + EX_2 + \ldots + EX_n \\ &= n\mu \\ Variance &= n\sigma^2 \end{aligned}$$

(b)

$$\begin{aligned} \bar{X} &= \frac{1}{n}(X_1 + X_2 + ... + X_n) \\ \text{Mean} &= \frac{1}{n}n\mu = \mu \\ \text{Variance} &= (\frac{1}{n})^2 n\sigma^2 = \sigma^2/n \end{aligned}$$

This implies that Z has the mean 0 and variance 1.

Derivation of equation (3.1)

Sampling from a normal distribution gives independent sample value. So using theorem (3.1), we can choose γ and then determine c such that

$$P(-c \leq Z \leq c) = P(-c \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c) = \phi(c) - \phi(-c) = \gamma$$

For the value $\gamma = 0.95$, we can find $z(D) = 1.960$.(Table A8,APP.5)(Erwin Kreyszig(2006)) $(D(z) = \phi(z) - \phi(-z))$

Now,

$$P(-c \leq Z \leq c) \quad = \quad P(c \geq -Z \geq -c)$$

$$
\begin{aligned}
&= P(c \geq \frac{\mu - \bar{X}}{\sigma/\sqrt{n}} \geq -c) \\
&= P(\frac{c\sigma}{\sqrt{n}} \geq \mu - \bar{X} \geq \frac{-c\sigma}{\sqrt{n}}) \\
&= P(K \geq \mu - \bar{X} \geq -K) \\
&= P(K + \bar{X} \geq \mu \geq -K + \bar{X}) \\
&= P(\bar{X} - K \leq \mu \leq \bar{X} + K) \\
&= \gamma.
\end{aligned}
$$

Inserting the observed values $\bar{x}$ of $\bar{X}$ gives (3.2). Hence we have regarded $x_1, x_2, \ldots, x_n$ as single observation of $X_1, X_2, \ldots, X_n$, so that $x_1 + x_2 + \ldots + x_n$ is an observed value of $X_1 + X_2 + \ldots + X_n$ and $\bar{x}$ is an observed values of $\bar{X}$. Here $\Theta_1 = \bar{X} - K$, and $\Theta_2 = \bar{X} + K$.

3.1.2 Confidence interval for mean μ of the Normal distribution with unknown variance σ^2

Here c depends on the sample size n, which determined form Table A9 in APP.5(Erwin Kreyszig(2006)). That table lists values z for given values of the distribution function,

$$F(z) = K_m \int_{-\infty}^{z} (1 + \frac{u^2}{m})^{-(m+1)/2} du$$

of the t-distribution.

Here m(=1,2,...) is a parameter, called the number of degrees of freedom of the distribution. In the present case, $m = n - 1$. The constant K_m is such that $F(\infty) = 1$. By integrating it, it will be

$$K_m = \Gamma(\frac{1}{2}m + \frac{1}{2})/[\sqrt{m\pi}\Gamma(\frac{1}{2}m)],$$

where Γ is a gamma function.

Determination of a confidence interval for mean μ of the normal distribution with unknown variance σ^2

*Step*1 : Choose a confidence level $\gamma(95\%, 99\%, \ldots)$.

*Step*2 : Determine the solution c of the equation

$$F(c) = \frac{1}{2}(1 + \gamma)$$

from the table of the t-distribution with $n-1$ degree of freedom.
*Step*3 : Compute the mean $\bar{x}$ and the variance s^2 of the sample $x_1, x_2, \ldots, x_n$.
*Step*4 : Compute $k = cs/\sqrt{n}$. The confidence interval is

$$CONF_{\gamma}\{\bar{x} - k \le \mu \le \bar{x} + k\}. \tag{3.4}$$

Example 3.2 *Five independent measurements of the point of inflammation (flash point) of Diesel oil (D-2) gave the values* (*in*^0F) 144 147 146 142 144. *Assuming normality determine a 99% confidence interval for the mean.*
Solution
*Step*1 : *Here* γ*=0.99 is required.*
*Step*2 : $F(c)$=$\frac{1}{2}(1+\gamma) = \frac{1}{2}(1.99) = 0.995$ *(Table A9 in APP.5)(Erwin Kreyszig(2006)),*
So we get $n-1=4$ *d.f which gives* $c = 4.60$
*Step*3 : $\bar{x} = \frac{144+147+146+142+144}{5} = 144.6$*,* $s^2 = 3.8$
*Step*4 : $k = c\sigma/\sqrt{n}$= $4.60\ \sqrt{3.8}/\sqrt{5} = 4.01$
The confidence interval is

$$CONF_{0.99}(\bar{x} - k \le \mu \le \bar{x} + k) = CONF_{0.99}(140.5 \le \mu \le 148.7).$$

t-Distribution Let $X_1, X_2, \ldots, X_n$ be independent normal random variable with the mean μ and the variance σ^2. Then the random variable

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

has t-distribution with n-1 degree of freedom, here $\bar{X}$ is given and,

$$S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2.$$

Derivation of equation (3.4)
We choose a number γ between 0 and 1 and determine a number c from Table A9 in App.5 (Erwin Kreyszig(2006)) with n-1 d.f. such that

$$P(-c \le T \le c) = F(c) - F(-c) = \gamma. \tag{3.5}$$

Since the t-distribution is symmetric, So we have,

$$F(-c) = 1 - F(c)$$
$$\begin{aligned} F(c) - (1 - F(c)) &= \gamma \\ \Rightarrow 2F(c) - 1 &= \gamma \\ \Rightarrow F(c) &= \frac{1}{2}(1+\gamma). \end{aligned}$$

Putting value of T in equation (3.5)

$$\begin{aligned} P(-c \leq T \leq c) = \gamma \\ \Rightarrow P(-c \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq c) = \gamma \\ P(\bar{X} - K \leq \mu \leq \bar{X} + K) = \gamma, \end{aligned}$$

where $K = cS/\sqrt{n}$.
Inserting the observed values $\bar{x}$ of $\bar{X}$ and s^2 of S^2, gives equation (3.4).

3.1.3 Confidence interval for the variance σ^2 of the Normal distribution

Determination of a Confidence Interval for the variance σ^2 of the Normal distribution, whose mean need not be known
$Step1$: Choose a confidence level $\gamma(95\%, 99\%, \ldots)$.
$Step2$: Determine the solution c_1 and c_1 of the equation

$$F(c_1) = \frac{1}{2}(1-\gamma),\ F(c_2) = \frac{1}{2}(1+\gamma)$$

from the table of the chi-square distribution with $n-1$ degree of freedom (Table A10 in App. 5)(Erwin Kreyszig(2006)).
$Step3$: Compute the $(n-1)s^2$, where s^2 is the variance of the sample $x_1, x_2, \ldots, x_n$
$Step4$: Compute $k_1 = (n-1)s^2/c_1$ and $k_2 = (n-1)s^2/c_2$. The confidence interval is

$$CONF_\gamma\{k_2 \leq \sigma^2 \leq k_1\}. \tag{3.6}$$

Example 3.3 *Determine a* 95% *confidence interval for the variance and a sample (tensile strength of sheet steel in* Kg/mm^2*, rounded to integer values)* 89 84 81 87 89 86 91 90 78 89 87 99 83 89.

Solution

*Step*1 : γ*=0.95 is required.*

*Step*2 : $n = 14$, $n - 1 = 13$, *we have*

$$\begin{aligned} F(c_1) &= \frac{1}{2}(1-\gamma) = \frac{1}{2}(1-0.95) = 0.025, \\ F(c_2) &= \frac{1}{2}(1+\gamma) = \frac{1}{2}(1+0.95) = 1.95 \end{aligned}$$

$\therefore c_1 = 5.01$ *and* $c_2 = 24.74$

*Step*3 : $(n-1)s^2 = 13s^2 = 326.9$

*Step*4 : $k_1 = (n-1)s^2/c_1$ *and* $k_2 = (n-1)s^2/c_2$, $k_1 = 65.25$ *and* $k_2 = 13.21$

The confidence interval is

$$CONF_{0.95}\{13.21 \le \sigma^2 \le 65.25\}.$$

For this we shall use the χ^2-distribution.

Chi-Square Distribution The distribution function of the χ^2-distribution is

$$F(z) = 0 \text{ if } z < 0$$

and

$$F(z) = C_m \int_0^z e^{-u/2} u^{(m-2)/2} du \quad \text{if } z \ge 0.$$

The parameter m(=1,2,3,...) is called the number of degree of freedom and

$$C_m = 1/\left[2^{m/2}\Gamma(\frac{1}{2}m)\right]$$

Here the distribution is not symmetric .

Here the random variable

$$Y = (n-1)S^2/\sigma^2$$

with $S^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$ has a chi-square distribution with $n-1$ degree of freedom.

Derivation of equation (3.6)
We choose a number γ between 0 and 1 and determine c_1 and c_2 from the table A10, App.5(Erwin Kreyszig(2006) such that,

$$
\begin{aligned}
P(Y \le c_1) &= F(c_1) = \frac{1}{2}(1-\gamma), \\
P(Y \le c_2) &= F(c_2) = \frac{1}{2}(1+\gamma). \\
P(c_1 \le Y \le c_2) &= P(Y \le c_2) - P(Y \le c_1) = F(c_2) - F(c_1) = \gamma.
\end{aligned}
$$

Putting values of Y from chi-square distribution

$$c_1 \le Y \le c_2$$

$$\Rightarrow \frac{n-1}{c_2}S^2 \le \sigma^2 \le \frac{n-1}{c_1}S^2.$$

By inserting the observed value s^2 of S^2 we obtain equation (3.6).

3.2 Binomial confidence intervals

A random variable X is a Bernoulli trial if it has two points in its sample space, That is $P(X=1)=p$, and $P(X=0)=1-p$. The expected value of X is $E[X]=p$, and the variance of X is $Var[X]=p(1-p)$.
Suppose $X_1, X_2, \ldots, X_n$ are a sample of size n from a Bernoulli distribution with parameter p. The sum $Y = X_1 + X_2 + \ldots X_n = \sum_j X_j$ is a random variable with sample space $0, 1, 2, \ldots, n$. It follows a binomial distribution.The probability mass function of Y is

$$P(Y=k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

The expected value of Y is $E[Y]=np$ and the variance is $Var[Y]=np(1-p)$.
We can estimate p using $\hat{p} = (X_1 + X_2 + \ldots + X_n)/n = Y/n$. This estimate is unbiased since $E[\hat{p}] = p$. The variance is $Var[\hat{p}] = p(1-p)/n$, and the standard deviation is $SD[\hat{p}] = \sqrt{p(1-p)/n}$. We can write $\hat{p} \sim [p, p(1-p)/n]$.
We can standardize any random variable by subtracting its mean and dividing the result

by the standard deviation. The resulting random variable has expected value 0 and variance 1

$$\frac{\hat{p}-p}{\sqrt{p(1-p)/n}} = \sqrt{n}\frac{\hat{p}-p}{\sqrt{p(1-p)}} \sim [0,1].$$

The normal theory approximation of a confidence interval for a proportion is known as the Wald interval, defined as $\hat{\pi} \pm z\sqrt{\hat{\pi}(1-\hat{\pi})/n}$, where z is the $z_{1-\alpha/2}$ quantile of the standard normal distribution. It has been known that Wald interval performs very good when n is large. Due to the discreteness of the binomial distribution the empirical value of the confidence coefficient is larger than the nominal level $1-\alpha$.
Wald type confidence interval for p

$$\begin{aligned}
&= P(-1.96 \le \sqrt{n}(\hat{p}-p)/\sqrt{p(1-p)}) \le 1.96) \\
&= P(-1.96/\sqrt{n} \le (\hat{p}-p)/\sqrt{p(1-p)}) \le 1.96/\sqrt{n}) \\
&= P(-1.96\sqrt{p(1-p)}/\sqrt{n} \le (\hat{p}-p)) \le 1.96\sqrt{p(1-p)}/\sqrt{n}) \\
&= P(-1.96\sqrt{p(1-p)}/\sqrt{n} - \hat{p} \le -p \le 1.96\sqrt{p(1-p)}/\sqrt{n} - \hat{p}) \\
&= P(\hat{p} - 1.96\sqrt{p(1-p)}/\sqrt{n} \le p \le 1.96\sqrt{p(1-p)}/\sqrt{n} + \hat{p}).
\end{aligned}$$

This gives us the Wald-type 95% confidence interval

$$\hat{p} \pm 1.96\sqrt{p(1-p)/n} \approx \hat{p} \pm 2\sqrt{p(1-p)/n}.$$

The lower confidence limit (LCL) is $\hat{p} - 1.96\sqrt{p(1-p)/n}$, and the upper confidence limit (UCL) is $1.96\sqrt{p(1-p)}/\sqrt{n} + \hat{p}$.

Chapter 4

Bayesian Interval Estimation

4.1 Exponential Distributions

Consider the one-parameter exponential probability density function

$$f(x|\theta) = \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right), \; x, \theta > 0.$$

The likelihood function

$$L(\underline{x}|\theta) = \frac{1}{\theta^n} \exp\left(-\frac{S}{\theta}\right), \quad S = \sum_{i=1}^{n} x_i,$$

Using Jeffreys' prior $g(\theta) \propto \frac{1}{\theta}$, the posterior distribution is

$$\Pi(\theta|\underline{x}) = \frac{k}{\theta^{n+1}} \exp\left(-\frac{S}{\theta}\right)$$

$$\begin{aligned}
C^{-1} &= \int_0^\infty \Pi(\theta|\underline{x}) d\theta \\
&= \int_0^\infty \frac{1}{\theta^{n+1}} \; \exp\left(-\frac{S}{\theta}\right) d\theta \\
&= \int_0^\infty \left(\frac{1}{\theta}\right)^{n+1} \exp\left(-\frac{S}{\theta}\right) d\theta
\end{aligned}$$

$$
\begin{aligned}
&= \int_0^\infty \left(\frac{S}{\theta}\right)^{n+1} \frac{1}{S^{n+1}} \exp\left(-\frac{S}{\theta}\right) d\theta.
\end{aligned}
$$

$$
\begin{aligned}
\text{Let } \frac{S}{\theta} &= t \Rightarrow -S\theta^{-2} d\theta = dt \Rightarrow d\theta = -\frac{S}{t^2} dt \\
&= \int_0^\infty t^{n+1} \frac{1}{S^{n+1}} e^{-t} \left(-\frac{S}{t^2}\right) dt \\
&= \frac{-S}{S^{n+1}} \int_0^\infty t^{n-1} e^{-t} dt \\
&= \frac{\Gamma n}{S^n} \\
\Rightarrow C &= \frac{S^n}{\Gamma n} \\
\Pi(\theta|\underline{x}) &= C g(\theta) f(x|\theta) \\
&= \frac{S^n}{\Gamma n} \frac{1}{\theta^{n+1}} e^{\frac{-S}{\theta}}, \quad \text{where } S = \sum_{i=1}^n x_i
\end{aligned}
$$

$$
\begin{aligned}
\chi^2(2n) &= \frac{x^{n-1} e^{\frac{-x}{2}}}{2^n \Gamma(n)} \ \text{ if } \ x \geq 0 \\
&= 0 \ \text{ if } \ x \leq 0.
\end{aligned}
$$

It follows that $\frac{2S}{\theta} \sim \chi^2(2n)$.

$$
\begin{aligned}
1 - \alpha &= P[\chi_1^2 < \chi^2 < \chi_2^2] \\
&= P[\chi_1^2 < \frac{2S}{\theta} < \chi_2^2] \\
&= P[\frac{2S}{\chi_2^2} < \theta < \frac{2S}{\chi_1^2}]
\end{aligned}
$$

$$\text{where } \chi_1^2 = \chi^2(1-\alpha/2, 2n)$$

$$\chi_2^2 = \chi^2(\alpha/2, 2n).$$

Let $[C_L^{(\theta)}, C_U^{(\theta)}]$ and $[H_L^{(\theta)}, H_U^{(\theta)}]$ respectively represent the lower and upper credible and HPD interval of the parameter θ. Thus the $100(1-\alpha)\%$ equal-tail credible interval for the exponential parameter θ is given by,

$$[C_L^{(\theta)}, C_U^{(\theta)}] = \left[\frac{2S}{\chi^2(\alpha/2, 2n)}, \frac{2S}{\chi^2(1-\alpha/2, 2n)}\right], \tag{4.1}$$

which is the same as the classical interval.
The posterior distribution of θ is unimodel. so the HPD interval of θ should simultaneously satisfy

$$1-\alpha = \int_{H_L^{(\theta)}}^{H_U^{(\theta)}} \Pi(\theta|\underline{x})dx$$
$$\text{and } \Pi(H_L^{(\theta)}|\underline{x}) = \Pi(H_U^{(\theta)}|\underline{x}).$$

So,We have

$$1-\alpha = P[H_L^{(\theta)} < \theta < H_U^{(\theta)}] = P\left[\frac{2S}{H_U^{(\theta)}} < \theta < \frac{2S}{H_L^{(\theta)}}\right] \tag{4.2}$$

$$\text{and} \quad \exp\left[-S\left(\frac{1}{H_L^{(\theta)}} - \frac{1}{H_U^{(\theta)}}\right)\right] = \left[\frac{H_L^{(\theta)}}{H_U^{(\theta)}}\right]^{n+1} \tag{4.3}$$

which are simultaneously satisfied.

Example 4.1 *Suppose we have a random sample $\{x_i\}, i = 1, 2, \ldots, 15$ from the one parameter exponential distribution we will obtain 95% credible and HPD interval for θ.*

Data: $\{x_i\}$
0.5142 7.8446 8.3758 10.5329 13.9725 3.9769 3.9779 4.3163 4.4317 4.7371
0.6454 0.8451 1.0566 1.2681 15.3301

$$S = \sum_{i=1}^{15} x_i = 81.8252$$

We have 95% *credible interval, So* $1-\alpha = 0.95$, α *=1- 0.95= 0.05*
From equation (4.1)

$$[C_L^{(\theta)}, C_U^{(\theta)}] = \left[\frac{163.6504}{\chi^2(0.025,30)}, \frac{163.6504}{\chi^2(0.975,30)}\right]$$

$$= [3.4835, 9.7470]$$

$$C_U^{(\theta)} - C_L^{(\theta)} = 6.2635$$

From equation (4.2) *and* (4.3)

$$1-\alpha = P\left[\frac{2S}{H_U^{(\theta)}} \quad < \chi^2(2n) \quad < \frac{2S}{H_L^{(\theta)}}\right]$$

$$0.95 = P\left[\frac{2(81.8252)}{H_U^{(\theta)}} \quad < \chi^2(2n) \quad < \frac{2(81.8252)}{H_L^{(\theta)}}\right]$$

$$= P\left[\frac{1}{H_U^{(\theta)}} \quad < \frac{0.0000004}{163.6504} \quad < \frac{1}{H_L^{(\theta)}}\right]$$

$$\Rightarrow P\left[H_L^{(\theta)} \quad < 409126000 \quad < H_U^{(\theta)}\right] = 0.95$$

$$\Rightarrow [H_L^{(\theta)}, H_U^{(\theta)}] = [3.2450, 9.1770]$$

$$H_U^{(\theta)} - H_L^{(\theta)} = 5.9320 \quad < \quad C_U^{(\theta)} - C_L^{(\theta)}.$$

4.2 Normal Distribution

Consider the Normal probability density function

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\ exp\ \{-\frac{1}{2\sigma^2}(x-\mu)^2\}, \quad -\infty < \mu, \quad \mu < \infty, \quad \sigma > 0$$

The likelihood function

$$L(\mu,\sigma|\underline{x}) = \frac{1}{(\sqrt{2\pi})^n\sigma^n}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\}$$

The marginal posteriors of μ and σ^2 with respect to the prior $g(\mu,\sigma^2)\propto\frac{1}{\sigma}$
The joint prior distribution of μ and σ is

$$\Pi(\mu,\sigma|\underline{x}) = \frac{k}{(\sqrt{2\pi})^n\sigma^n}\exp\left[-\frac{1}{2\sigma^2}\{A+n(\bar{x}-\mu)^2\}\right],\quad A=\sum_{i=1}^{n}(x_i-\bar{x})^2$$

where value of k is given below

$$\begin{aligned}
k^{-1} &= \frac{1}{\sigma\sqrt{2\pi}}\int_0^\infty\frac{\exp(-A/2\sigma^2)}{\sigma^n}\left[\int_0^\infty\exp\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\}d\mu\right]d\sigma\\
&= \frac{\sqrt{2\pi}}{n}\int_0^\infty\frac{\exp(-A/2\sigma^2)}{\sigma^{n-1}}d\sigma\\
&= \sqrt{\frac{\pi}{2n}}\int_0^\infty\frac{\exp(-A/2\sigma^2)}{(\sigma^2)^{n-1/2}}d\sigma\\
&= \sqrt{\frac{\pi}{2n}}\Gamma(\frac{n-1}{2})(\frac{2}{A})^{n-1/2}
\end{aligned}$$

Restoring the constant k and integrating out σ from $\Pi(\mu,\sigma|\underline{x})$, we have marginal posteriors of μ.

$$\Pi_1(\mu|\underline{x}) = \frac{\sqrt{n/A}}{B(\frac{1}{2},\frac{n-1}{2})(1+\frac{n(\bar{x}-\mu)^2}{A})^{(n-1)/2}} \tag{4.4}$$

Similarly, we obtain the marginal posterior of σ^2

$$\begin{aligned}
\Pi(\sigma^2|\underline{x}) &= \int_{-\infty}^{\infty}\pi(\mu,\sigma|\underline{x})d\mu\\
&= k\sqrt{\frac{\pi}{2n}}\frac{\exp(-A/2\sigma^2)}{(\sigma^2)^{n/2}}
\end{aligned}$$

$$= \frac{(A/2)^{n-1/2}}{\Gamma(\frac{n-1}{2})}\frac{\exp(-A/2\sigma^2)}{(\sigma^2)^{n+1/2}},\quad 0<\sigma<\infty \tag{4.5}$$

From equation (4.4) it follows that $\frac{\sqrt{n-1}(\mu-\bar{x})}{\sqrt{A/n}}\sim$ Student's t-distribution with $(n-1)$ degree of freedom.

Also, $\Pi(\mu|\underline{x})$ is unimodal and symmetric about the mean $\bar{x}$. Hence, the $(1-\alpha)$ equal-tail, the shortest credible and HDP interval for μ are identical such an interval (L,U)

must satisfy the condition,

$$
\begin{aligned}
1-\alpha &= P(L<\mu<U) \\
&= P\left[\frac{\sqrt{n-1}(L-\bar{x})}{\sqrt{A/n}} < t < \frac{\sqrt{n-1}(U-\bar{x})}{\sqrt{A/n}}\right] \\
&= P\left[\frac{(L-\bar{x})}{S/\sqrt{n}} < t < \frac{(U-\bar{x})}{S/\sqrt{n}}\right] \\
&= P\left[-t(\alpha/2, n-1) < t < t(\alpha/2, n-1)\right]
\end{aligned}
$$

Hence, we have

$$
L = \bar{x} - \frac{S}{\sqrt{n}} t(\alpha/2, n-1), \quad U = \bar{x} + \frac{S}{\sqrt{n}} t(\alpha/2, n-1), \tag{4.6}
$$

where $t(\alpha, \nu) = 100\alpha\%$ point of student's t-distribution with ν degree of freedom.

The $(1-\alpha)$ -HDP interval of μ is the same as the classical confidence counterpart.

From (4.5) it follows that $\frac{A}{\sigma^2}$ is distribution as a χ^2 with (n-1) degree of freedom. The $(1-\alpha)$ - equaltail credible interval $[T_1, T_2]$ must satisfy,

$$
\alpha/2 = \int_0^{T_1} \Pi(\sigma^2/\underline{x}) d\sigma^2 = \int_{T_2}^{\infty} \Pi(\sigma^2/\underline{x}) d\sigma^2
$$

$$
\begin{aligned}
\alpha/2 &= P(\sigma^2 \le T_1) \\
&= P\left[A/\sigma^2 \ge A/T_1\right] \\
&= P\left[\chi^2 \ge A/T_1 = \chi^2(\alpha/2, n-1)\right]
\end{aligned}
$$

Similarly,

$$
1-\alpha/2 = P\left[\chi^2 \le A/T_1 = \chi^2(1-\alpha/2, n-1)\right]
$$

Thus, the $(1-\alpha)$- equal tail credible interval of σ^2 is

$$
[T_1, T_2] = \left[\frac{A}{\chi^2(\alpha/2, n-1)}, \frac{A}{\chi^2(1-\alpha/2, n-1)}\right] \tag{4.7}
$$

where $\chi^2(\alpha, \nu)$= upper 100% point of a χ^2-distribution with ν degree of freedom.

4.3 Weibull Distribution

Consider the Weibull density function

$$f(x|\ \theta, P) = \frac{P}{\theta}\ x^{P-1}\exp(-x^P/\theta), \quad P, \theta, x > 0$$

The likelihood function

$$L(\theta|\underline{x}) = P^n\theta^{-n}\lambda^{P-1}\exp(-\sum_{i=1}^{n} x_i^P/\theta), \quad \lambda = \Pi_{i=1}^{n}x_i$$

Consider jeffreys prior $h(\theta, P) \propto \frac{1}{\theta P}$. We obtain the joint posterior of (θ, P)

$$\Pi(\theta, P|\ \underline{x}) = kP^{n-1}\theta^{-(n+1)}\lambda^{P-1}\exp(-\sum_{i=1}^{n} x_i^P/\theta), \tag{4.8}$$

where

$$\begin{aligned} k_1^{-1} &= \int_0^\infty\int_0^\infty \Pi(\theta, P|\underline{x})d\theta\ dP \\ &= \int_0^\infty P^{n-1}\left[\int_0^\infty \theta^{-(n+1)}\exp(-\sum_{i=1}^{n} x_i^P/\theta)d\theta\right] dP \\ &= \Gamma(n)\int_0^\infty \frac{P^{n-1}\lambda^{P-1}}{\{\sum_{i=1}^n x_i^P\}^n}\ d\theta \end{aligned}$$

Thus,

$$\Pi(\theta, P|\ \underline{x}) = \frac{P^{n-1}\theta^{-(n+1)}\lambda^{P-1}\exp(-\sum_{i=1}^n x_i^P/\theta)}{\Gamma(n)\int_0^\infty \frac{P^{n-1}\lambda^{P-1}}{\{\sum_{i=1}^n x_i^P\}^n}} \tag{4.9}$$

Similarly, integrating out P from (4.9) the marginal posterior of θ is given by

$$\Pi_2(\theta|\underline{x}) = k_2\theta^{-(n+1)}\int_0^\infty P^{n-1}\lambda^{P-1}\exp(-\sum_{i=1}^{n} x_i^P/\theta)\ dP, \tag{4.10}$$

where $k_2 = \Gamma(n)/k_1$

The $(1-\alpha)$-crdible interval for P is given by $[C_1^{(P)}, C_1^{(P)}] = [P_1, P_2]$, where

$$\alpha/2 = \int_0^{P_1} \Pi_1(P|\ \underline{x})dP = \int_{P_2}^\infty \Pi_2(P|\ \underline{x})\ dP$$

If $\Pi_1(P|\ \underline{x})$ is unimodal, the corresponding HPD interval $[H_1^{(P)}, H_1^{(P)}] = [H_1, H_2]$, where H_1 and H_2 Simultaneous satisfying

$$\Pi_1(H_1|\ \underline{x}) = \Pi_2(H_2|\ \underline{x})$$

and

$$\int_{H_1}^{H_2} \Pi_1(\theta|\ \underline{x}) dP = 1 - \alpha,$$

where,$\Pi_2(\theta|\underline{x})$ is unimodal.
Similarly, for the parameter θ, the $(1-\alpha)$-credible interval for θ is given by $[C_1^{(\theta)}, C_1^{(\theta)}] = [\theta_1, \theta_2]$, where

$$\int_{0}^{\theta_1} \Pi_2(\theta|\ \underline{x}) dx = \int_{\theta_2}^{\infty} \Pi_2(\theta|\ \underline{x}) d\theta = \alpha/2,$$

and the corresponding HDP interval $[H_1^{(\theta)}, H_1^{(\theta)}] = [h_1, h_2]$, where $\Pi_2(h_1|\ \underline{x}) = \Pi_2(h_2|\ \underline{x})$ and

$$\int_{h_1}^{h_2} \Pi_2(\theta|\ \underline{x}) d\theta = 1 - \alpha.$$

Conclusions and Scope of Future Works

In this project work, at first the basic requirements for the development of subsequent chapters are studied. Further some techniques for finding confidence interval of parameters of a distribution function are discussed. In particular the problem has been considered for the case of normal and binomial distribution. Also we have studied the confidence interval of parameters of exponential, normal and finally weibull distribution using Bayesian approach. Some of the future works to be carried out are listed below.

- Some other techniques can be derived for finding confidence interval of parameters of some non-normal distribution function.
- A review work can be done on testing of hypothesis regarding the parameters of a distribution.
- A simulation study can be done to compare all the estimators provided in the project work.

Bibliography

Clarke G. M. and Cooke D. (1983): A Basic Course in Statistics, Edward Arnold Ltd.,Australia.

Dougherty Edward R. (1990): Probability and Statistics for the Engineering, Computing and Physical Science, Prentice Hall, Englewood Clifts, New Jersey 07632.

Kreyszig Erwin (2006): Advance Engineering Mathematics, Ohio State University, Columbus, Ohio

Medhi J. (2006): Statistical Methods An Introductory Text, New Age International Publishers.

Miller Irwin and Miller Marylees (2009): John E.Freund's Mathematical Statistics, Prentice-Hall of India, New Delhi.

Rohatgi, V. K. and Saleh, A. K. Md. E.(2009): An Introduction to Probability and Statistic, Wiley-interscience, New York.

Shedden Kerby (2013): Binomial confidence intervals, Department of Statistics, University of Michigan

Sinha, S. K. (2004): Bayesian Estimation, New Age International, New Delhi.